U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

MESTRADO INTEGRADO EM MEDICINA

Instituto Ciências Biomédicas Abel Salazar
Universidade do Porto
2013/2014

Analysis of a novel e-learning system in measuring medical students outcomes in study time and recall accuracy

RUI ALEXANDRE DO PRADO COSTA

U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Rui Alexandre do Prado Costa

Analysis of a novel e-learning system in measuring medical students outcomes in study time and recall accuracy

Mestrado Integrado em Medicina

Área: Educação Médica

Analysis of a novel e-learning system in measuring medical students outcomes in study time and recall accuracy

Tiago Taveira-Gomes*[1,2]
Rui Prado Costa*[2,3]
Milton Severo[1]
Maria Amélia Ferreira[1]

[1]Center for Medical Education, Faculty of Medicine of the University of Porto, PT
[2]ALERT Life Sciences Computing SA
[3]Abel Salazar Biomedical Sciences Institute, University of Porto
Email: Tiago Taveira-Gomes*- tiago.taveira@me.com; Rui Prado Costa – rcosta4540@gmail.com
*Corresponding authors

Abstract

Background: Spaced-repetition and repeated testing with feedback, are two methodologies known to boost knowledge retention. ALERT STUDENT, an online platform for medical education, was devised to implement both approaches in an integrated fashion and promote research in this field, by allowing creation and distribution of learning objects named *flashcards*. This study aims to demonstrate how learning objects and repeated testing using open-ended questions can be brought together to provide insights on recall accuracy, and to characterize the extent to which student factors, content factors and repetition affect it.

Methods: Three in-person sessions (s0, s1 and s2), separated by one-week intervals were conducted with 96 medical students using the platform. Students were randomized into experiment and control groups using simple randomization. The study consisted on a 20 min study task using 27 *flashcards* about the Golgi complex, and on a self-assessment task, that consisted on answering *flashcard* open-ended questions and grading recall accuracy. On s0 both groups performed the self-assessment task. On s1 and s2, the experiment group performed the study task followed by the self-assessment task, whereas the control group only performed the self-assessment task.

Results: At s0 there were no differences in recall accuracy between groups, but in s1 and s2 recall accuracy differences were statistically significant. The experiment group achieved a sharper increase in recall accuracy than the control group, which was stronger between s0 and s1. In the experiment group, increases in recall accuracy were mainly due to the session effect (49.6%) followed by *flashcard* factors (15.3%) and student factors (5.1%). In the control group, increases in recall accuracy were mainly accounted by *flashcard* factors (34.7%), followed by student (15.9%) and session (8.2%) factors. Generalizability score for the *flashcard* in the control group was 0.91, indicating almost perfect agreement, and 0.47 for the experiment group, indicating moderate agreement. This finding indicates that *flashcards* can be very well characterized by recall accuracy scores by having the same students perform sequential self-assessment tasks alone. Regarding the experiment group, differences in study duration for sessions s1 and s2 were statistically significant. Mean study duration was 14:32 minutes for s1 and 12:28 minutes for s2.

Conclusions: The self-assessment activity alone had little effect on recall accuracy, however studying and self-test activities together greatly increased recall accuracy. The session effect was the main determinant of recall accuracy for the experiment group. Learning objects may be able to provide additional value when combined with open-ended questions. Insights about recall accuracy can be useful for teachers to assess content difficulty, learning rate, and for students to prioritize content and make better use of study sessions.

Keywords

Medical education, Memory retention, Computer-Assisted Instruction, E-learning, Tailored-learning, Spaced repetition

Background

Medical education is a complex field where updates in medical knowledge, educational technology and teaching strategies intertwine in a progressive fashion [1, 2, 3, 4, 5]. Over the past decade there has been a shift in this field, where traditional instructor-centered teaching is yielding to a learner-centered model [6, 7, 8, 9], in which the learner has greater control over the learning methodology and the role of a teacher becomes that of a facilitator of knowledge acquisition, replacing the role of an information provider [10, 11, 12, 7].

There is vast literature regarding the application of educational strategies [13, 14, 15, 16, 17, 18], instructional design [19, 11, 20, 21, 22, 23] and cognitive learning science [24, 25, 26, 27] to the field of medical education. Two approaches that emerge from that literature are spaced-repetition and repeated testing.

Spaced-repetition is a learning approach grounded in cognitive learning theory that enforces continuously revisiting content over optimized time intervals, and has been shown to improve long-term knowledge retention [28, 29, 30, 11, 31].

Repeated testing during learning is an approach that extends spaced repetition by applying it to assessments during learning periods. It argues that the addition of repeated assessment feedback can further enhance learning [32, 33, 34]. In addition, the usage of open-ended questions as means of assessment has been shown to be superior to multiple choice questions in terms of knowledge retention [35].

Despite this vast literature, there are few reports of implementing these principles in an on-line platform in an integrated fashion [36].

ALERT STUDENT, a platform for medical education, was devised with the goal to implement such principles and promote research in this field [36]. The platform allows creation and distribution of learning objects [37] named *notebooks*. These are composed of smaller learning objects named *flashcards*, self-contained information chunks with related open-ended questions [36]. *Notebooks* can be accessed using a *study mode* that presents the *flashcards* in a study-friendly environment enriched with note taking, text highlighting and time tracking features. The *quiz mode* is a complementary environment where *flashcard* knowledge can be self-assessed using its open-ended questions. A set of questions from each *flashcard* is sequentially presented and requires the student to actively recall the answer and grade recall

accuracy using a 4 point *likert* scale. This information is summarized into a 4 point color-coded score that is attributed to each *flashcard* and displayed in the *study mode* as a study priority cue.

Exploring the usefulness of providing knowledge retention tools in e-learning systems is an important contribution to information management axis of the core-competences for medical education [38] and may provide a means to a richer and faster acquisition of factual knowledge, which is a cornerstone for the development of cognitive schemata, clinical reasoning skills [27].

This study aims to demonstrate how learning objects and repeated testing using open-ended questions can be brought together to provide insights on recall accuracy, and to characterize the extent to which student factors, content factors and repetition affect it.

Methods

Study design

The Faculty of Medicine of the University of *Porto* (FMUP) implements a 6-year graduate program. Applicants are mainly high school graduates. The first three years focus on basic sciences while the last three focus on clinical specialties.

Ninety-eight (n=98) students from the 4th and 5th grades of our school were contacted via email and volunteered to participate in this study. The study was performed in three in-person sessions (s0, s1 and s2) of 1 hour duration. Sessions were carried one week apart. Students were assigned into experiment group or control group using simple randomization.

The study consisted of a study task and a self-assessment task. During the study task, participants studied a *notebook* containing *flashcards* about the Golgi complex during 20 minutes using the *study mode*. During the self-assessment task, students were presented open-ended questions from the *notebook* on the *quiz mode*, and were required to grade their active recall for each question using a 4 point *likert* scale (0 - no recall, 1 - scarce recall, 2 - good recall, 3 - full recall). This task lasted 15 minutes. (*Study mode* and *quiz mode* environment screenshots can be viewed on appendix C)

A pilot study was conducted using two 2nd grade students that had recently completed the Cellular and Molecular Biology class, two 4th grade and two 5th grade medical students (n = 6). Students performed the self-assessment task. The 2nd grade students' average recall accuracy was assumed to be similar to what would be found at session s2. The 4th and 5th grade students' mean recall accuracy was assumed to correspond to what would be found at session s0. Based on the differences between the two pilot groups it was concluded that n = 48 would be sufficient to discriminate statistical significant differences in recall accuracy between sessions and between groups. The sample size was incremented to n = 98 to take advantage of the study venue capacity.

On s0, both groups performed the self-assessment task. On s1 and s2, the control group performed the self-assessment task, and the experiment group performed the study task, immediately followed by the self-assessment task.

For each session, *flashcard* recall accuracy was computed as the mean recall accuracy of the questions belonging to a *flashcard*. For the experiment group, study duration for each user and session was computed as the cumulative sum of time spent on each *flashcard* for the session.

Content design

The content used in this study was designed using official lecture content provided by the Cellular and Molecular Biology Department of our school, that is responsible for teaching the Golgi complex in Cellular Biology class as part of the first grade of the school medical curricula. See Appendix B to access notebook content. The questions were part of a *notebook* about the Golgi complex that contained 27 *flashcards*. In total 42 open ended questions about the 27 *flashcards* were answered and graded according to recall accuracy. The learning material did not change between sessions. During each session different questions were presented, but all of them covered the *notebook* material in a complete and homogeneous fashion.

Fitting content to time constraint

Another pilot study was performed to tailor the *notebook* content based on the study time constraint of 20 minutes. Another set of 6 5th grade students were assigned to read the *notebook* content. From the initial set of 30 *flashcards*, the last 3 were removed so that the content could be fully read within the 20 minute interval. No changes were necessary to answer the *quiz mode* questions within the 15 minute interval.

Sample characterization

In session s0 both groups filled a survey to characterize the student sample. Measured factors were gender, course year, preferred study resource for Cellular Biology, computer usage habits, Cellular Biology grade, mean course grade, and average study session duration during the semester and during the exam season.

Evaluation

Changes across sessions in recall accuracy were examined by univariate repeated-measures analysis of variance (ANOVA). Control and experiment groups were used as between-subjects factor. Session and *flashcard* were used as within subject factor. Repeated contrast (s0 vs s1 and s1 vs s2) was used to evaluate the sessions and the session interaction effect. In order to determine the reliability of the variance of recall accuracy explained by the *flashcard*, the

generalizability score (G-Score) for this variance component was computed. Guidelines for interpreting G-Scores suggest that values for relative variance between 0.81 - 1.00 indicate almost perfect agreement, 0.61 - 0.80 substantial agreement, 0.41 - 0.60 moderate agreement, 0.21 - 0.40 fair agreement, and values less than 0.21 depict poor or slight agreement [39]. The Wilcoxon signed-rank test was performed to compare study duration distributions of the experiment group for sessions s1 and s2.

This study was approved by the Faculty of Medicine University of *Porto* / *São João* Hospital Ethics Committee in compliance with the Helsinki Declaration.

Results

Study sample characterization

96 participants completed the session s0. 1 participant in the experiment group and 1 participant in the control group did not complete session s1 and were excluded from the study. By the end of session s2 there were 47 participants in each group. 59 participants were female and 35 participants were male. 44 participants were enrolled in the 4th grade and 53 were enrolled on the 5th grade. The preferred study resource for Cellular Biology was the *Professor texts* (n=36), followed by *Lecture notes* (n=24), *Lecture slides* (n=23) and finally the *Textbook* (n=11). Most participants reported using computers every day (n=78). Regarding average grades, average course grade was 12.8/20, and the average Cellular Biology grade was 13.6/20. Participants reported daily study sessions during the semester to last on average 3.0 hours and daily exam preparation study sessions to last on average 9.5 hours. No significant differences between the experiment group and control group distributions were found for any of the sample characterization factors. These results are described in further detail in Table A1.

Recall accuracy characterization

Mean recall accuracy increased from 0.76 in s0, to 1.59 in s1, to 1.87 in s2. In the control group, mean recall accuracy increased from 0.72 in s0 to 0.99 in s1 ($p < 0.001$) to 1.26 in s2 ($p < 0.001$). In the experiment group, recall accuracy increased from 0.81 at s0 to 2.18 at s1 ($p < 0.001$) to 2.47 at s2 ($p < 0.001$). At session s0, there were no differences in recall accuracy distributions between groups. During s1 and s2, recall accuracy differences between groups were statistically significant ($p < 0.001$). The experiment group achieved a sharper increase in recall accuracy than the control group. The increase in recall accuracy was greater between s0 and s1 for both groups. These results are described in further detail in Table A2.

Regarding the components of variance for recall accuracy in the control group, the largest one was the *flashcard* (34.7%). The *participant* and *session* components explained a small proportion of variance (15.1% and 8.2%, respectively) reflecting small systematic differences among participants and sessions. The residual component accounted for 41.2% of the variance. These results are described in further detail in Table A3.

In respect to the components of variance for recall accuracy in the experiment group, the most prominent factor was the *session* (49.6%). The *participant* and *flashcard* components explained

a small proportion of variance (5.1% and 15.3%, respectively). The residual component accounted for 30.0% of the variance. These results are described in further detail in Table A4.

For both groups two-way and three-way interactions were computed, but explained a very small fraction of total variance.

G-Score for *flashcard* variance in the control group was 0.91 , indicating almost perfect agreement, and for the experiment group was 0.47, indicating moderate agreement. To achieve an almost perfect agreement G-Score for the *flashcard* component (>0.80), having 5 students grade their recall accuracy on 3 separate sessions would be enough. Figures A1 describes *flashcard* difficulty G-Scores for different session and student count combinations, for control and experiment groups.

Study duration results for the experiment group

Median total study duration was 14:32 min in s1 and 12:28 min in s2. Standard deviations were 4:23 and 4:12 respectively. Mean difference between s1 and s2 study duration was approximately 2 min. Differences in total study duration were statistically significant ($p = 0.015$).

Discussion

Evolution of recall accuracy across sessions

There is an effect on recall accuracy reported by students along sessions. It was expected that the experiment group would out perform the control group in terms of recall accuracy, at least on s1. Since using open-ended assessment questions as a means to learn improves knowledge retention [33, 40, 41], it was unclear how strong would that increase be in the control group. However that increase was only a modest one. That finding might be explained, at least in part, by minimization of the cueing effect - the ability to answer questions correctly because of the presence of certain questions elements [42, 43] - through the usage of different variations of open-ended questions. Open-ended short-answer questions are known to minimize cueing [44, 43] and in addition, the variations in the question formulation for each session likely led to a stronger minimization of that effect. Furthermore, open ended questions are easier to design, which increases their overall utility as a self-assessment tool for systematic use.

The recall accuracy increase was stronger in the first session for the experiment group (63% gain). It was expected to see an increase in this session since the content was tailored to be fully covered within the 20 minute time limit. The strong gain indicates that this session was the one that accounted for the greatest increase in recall accuracy. On the second session there was still a significant but smaller increase in recall accuracy (12%). We would expect this effect to level off close to perfect recall (approximately 3.0) if sessions were to be repeated over time under the same circumstances.

Adequacy of recall accuracy as a measurement of knowledge

The sharp differences in recall accuracy between groups showed that this measurement, although being of subjective nature, seemed to be little affected by affective factors such as motivation and satisfaction. Even though it has been shown that in somewhat similar settings the relationship of knowledge self-assessment with motivation and satisfaction are stronger than with cognitive learning [45, 46, 47], such sharp differences between groups would not be expected, had these factors played major roles in the assessment. This finding leads to the hypothesis that performing the self-assessment immediately after the recall effort in addition to the feedback answer provided when grading may help students better understand their degree of knowledge and thus make sound assessments of their recall accuracy. However, this effect

can only be clearly ascertained by comparing the recall accuracy scores to objective knowledge measurements such as a multiple choice test, which is an aspect to investigate on further work.

Components of recall accuracy variance

Regarding the control group, the recall variance was mainly affected by *flashcard* features (34.7%) and by participant features (15.1%). This indicates, firstly, that the complexity of the content was mainly responsible for the recall scores across the sessions, and secondly, that differences in participants most likely regarding prior knowledge played a relevant role as well. The effect of the multiple sessions accounted little for the increase in recall accuracy over the sessions (8.2%). In addition, the experiment group had additional time to contact with the content, which by itself may also explain the differences found. The G-Score for the *flashcard* component was 0.91, indicating that *flashcards* can be very well characterized by recall accuracy scores by having the same students perform sequential self-assessment tasks without studying in between.

With respect to the experiment group, the contact with the content over multiple sessions was the main driver behind recall accuracy improvement (49.6%). Participant features such as prior knowledge had little effect in the increase in overall recall accuracy over sessions (5.1%), and the *flashcard* intrinsic difficulty also accounted for less effect than in the control group (15.1%). This suggests that the students in the experiment group increased their knowledge about the content and their prior knowledge had little effect in the learning process when using the study tools. These findings are inline with other studies that show that there is benefit in using repeated testing with study session in order to enhance learning [33, 40, 41].

Potential implications to teachers

The way in which content can be organized to optimize learning has been extensively studied [48, 49, 50, 37, 26, 38]. Learning objects, groupings of instructional materials structured to meet specific educational objectives [37], define a set of guidelines to make content portable, interactive and reusable, [51, 37, 9, 52, 53] therefore enhancing and tailoring learning [52]. They may facilitate adaptive learning by offering the chunks of content that the learner needs in order to achieve the required level of knowledge. This study demonstrates how learning objects can provide additional value when combined with open ended questions. By revealing

detailed insight about recall accuracy, teachers could use this information to classify course content based on student recall accuracy. By providing diagnostic tests on the beginning of a course, similar to the control group self-assessment task, teachers can quantify content complexity for each student class, and tailor content, lectures and other activities to help students meet the course goals. Provided that students study course contents, situations where the recall accuracy main source of variance is not the study session may alert teachers to flaws in content design, excessive course difficulty or inefficacies in teaching and learning methodologies. Cases where the increase in recall accuracy are mainly accounted by student features are possibly explained by strong discrepancies in student background knowledge. Regardless of the case, as long as students are committed to study, the desired is having study sessions explain the majority recall accuracy increases.

Potential implications to learners

Students need tools to help retain knowledge for longer periods and easily identify harder materials [31]. This goal may be achieved by providing learners with personal insight on their learning effectiveness, using personal and peer progress data based on self-assessment results [52]. The past recall accuracy can be used as a explicit cue to prioritize learning objectives and help better manage study time. Prioritizing content using explicit recall accuracy cues and grading after the study sessions are important tools to motivate students and optimize time studying. This is even more important at a time where students need to define tangible goals that allow them cope with course demands [54]. Additionally if content is distributed as learning objects using ALERT STUDENT, it can be segmented and reused in multiple steps of the academic curricula, and therefore recall accuracy scores acquired in different contexts can help manage and prioritize learning in new settings [36].

Limitations and further work

This work has several limitations. Although it can be hypothesized that the effect of these findings should hold or possibly increase in an ecological scenario, general applicability of these findings requires further research using different areas of medical curriculum, under larger learning workloads and realistic use scenarios. It is not known whether study and self-assessment sessions are more effective than study sessions only. Other authors have shown that study activities not followed by assessment activities result in less knowledge retention than study plus assessment activities [34, 40]. A limitation of this study is that it was not

possible to ascertain how much effect of recall accuracy is attributable to the study session and how much is attributable to the self-assessment session in the experiment group.

This study uses content that was previously taught to the students approximately 3 years before the study. These findings may not apply in scenarios where the students are exposed to content for the first time.

To our knowledge, no previous research has described recall accuracy in a controlled setting using an online e-learning platform in the medical curriculum. Future research questions regard understanding the relationship between recall accuracy and knowledge acquisition using objective assessment methods, and assessing the benefits of tailoring course design using recall accuracy information.

Conclusions

The present study focus on measuring recall accuracy of content using open ended questions in a controlled setting using the ALERT STUDENT platform. It was found that self-assessment recall activity alone led to a modest increase on recall accuracy, and that studying and self-assessment together had high impact in recall accuracy. It was shown that session repetition was the main determinant of recall accuracy when students perform study followed by the self-assessment task, and that when employing self-assessment only, content and student factors determine most of the increase in recall accuracy. These insights may be helpful to tailor teaching and learning methodologies to better reach learning goals.

The present findings will be explored in more detail in future work, as they may help future physicians and medical schools meet the challenge of information management [38] and instilling a culture of continuous learning, underpinning the core competencies outlined for XXI century physicians [4] [3].

List of abbreviations

ANOVA Analysis of variance
FMUP Faculty of Medicine, University of *Porto*
G-Score Generalizability score
s0 Study session 0
s1 Study session 1
s2 Study session 2

Competing interests

The authors of the study declare no competing interests.

Authors contributions

TTG and RC formulated the research problem, carried on-site procedures, performed statistical analysis and contributed equally to this work. MS Validated study design and conducted statistical analysis. MAF oversaw the study and gave final approval. All authors contributed to the study design and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors wish to thank to the students who took part in the study.

References

1. Qiao YQ, Shen J, Liang X, Ding S, Chen FY, Shao L, Zheng Q, Ran ZH: **Using cognitive theory to facilitate medical education**. *BMC medical education* 2014, **14**:79.

2. Schoonheim M, Heyden R, Wiecha JM: **Use of a virtual world computer environment for international distance education: lessons from a pilot project using Second Life**. *BMC medical education* 2014, **14**:36.

3. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H: **Health professionals for a new century: transforming education to strengthen health systems in an interdependent world.** *Lancet* 2010, **376**(9756):1923–58, [http://www.ncbi.nlm.nih.gov/pubmed/21112623].

4. Horton R: **A new epoch for health professionals' education.** *Lancet* 2010, **376**(9756):1875–7, [http://www.ncbi.nlm.nih.gov/pubmed/21112621].

5. Patel VL, Cytryn KN, Shortliffe EH, Safran C: **The collaborative health care team: the role of individual and group expertise.** *Teaching and learning in medicine* 2000, **12**(3):117–32, [http://www.ncbi.nlm.nih.gov/pubmed/11228898].

6. Bahner DP, Adkins E, Patel N, Donley C, Nagel R, Kman NE: **How we use social media to supplement a novel curriculum in medical education.** *Medical teacher* 2012, [http://informahealthcare.com/doi/abs/10.3109/0142159X.2012.668245].

7. Ruiz JG, Mintzer MJ, Leipzig RM: **The impact of E-learning in medical education.** *Academic medicine : journal of the Association of American Medical Colleges* 2006, **81**(3):207–12, [http://www.ncbi.nlm.nih.gov/pubmed/16501260].

8. Eysenbach G: **Medicine 2.0: social networking, collaboration, participation, apomediation, and openness.** *Journal of medical Internet research* 2008, **10**(3):e22, [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2626430\&tool=pmcentrez\&rendertype=abstract].

9. Boulos MNK, Maramba I, Wheeler S: **Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education.** *BMC medical education* 2006, **6**:41, [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1564136\&tool=pmcentrez\&rendertype=abstract].

10. Koops W, Van der Vleuten C, De Leng B, Oei SG, Snoeckx L: **Computer-supported collaborative learning in the medical workplace: Students' experiences on formative peer feedback of a critical appraisal of a topic paper.** *Medical teacher* 2011, **33**(6):e318–23, [http://www.ncbi.nlm.nih.gov/pubmed/21609168].

11. Cook Da, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM: **Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis.** *Academic medicine : journal of the Association of American Medical Colleges* 2010, **85**(5):909–22, [http://www.ncbi.nlm.nih.gov/pubmed/20520049].

12. Chretien KC, Greysen SR, Chretien JP, Kind T: **Online posting of unprofessional content by medical students.** *JAMA : the journal of the American Medical Association* 2009, **302**(12):1309–15, [http://www.ncbi.nlm.nih.gov/pubmed/19773566].

13. Woltering V, Herrler A, Spitzer K, Spreckelsen C: **Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation.** *Advances in health sciences education : theory and practice* 2009, **14**(5):725–38, [http://www.ncbi.nlm.nih.gov/pubmed/19184497].

14. Sandars J, Haythornthwaite C: **New horizons for e-learning in medical education: ecological and Web 2.0 perspectives.** *Medical teacher* 2007, **29**(4):307–10, [http://www.ncbi.nlm.nih.gov/pubmed/17786742].

15. Giani U, Brascio G, Bruzzese D, Garzillo C, Vigilante S: **Emotional and cognitive information processing in web-based medical education.** *Journal of biomedical informatics* 2007, **40**(3):332–42, [http://www.ncbi.nlm.nih.gov/pubmed/17208055].

16. Pereira JA, Pleguezuelos E, Mer_ A, Molina-Ros A, Molina-Tom_s MC, Masdeu C: **Effectiveness of using blended learning strategies for teaching and learning human anatomy.** *Medical education* 2007, **41**(2):189–95, [http://www.ncbi.nlm.nih.gov/pubmed/17269953].

17. Ruiz JG, Mintzer MJ, Leipzig RM: **The impact of E-learning in medical education.** *Academic medicine : journal of the Association of American Medical Colleges* 2006, **81**(3):207–12, [http://www.ncbi.nlm.nih.gov/pubmed/16501260].

18. Taradi SK, Taradi M, Radic K, Pokrajac N: **Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology.**

Advances in physiology education 2005, **29**:35–9, [http://www.ncbi.nlm.nih.gov/pubmed/15718381].

19. Conn JJ, Lake FR, McColl GJ, Bilszta JLC, Woodward-Kron R: **Clinical teaching and learning: from theory and research to application**. *The Medical Journal of Australia* 2012, **196**(8):527, [http://www.ncbi.nlm.nih.gov/pubmed/22571313https://www.mja.com.au/journal/2012/196/8/clinical-teaching-and-learning-theory-and-research-application].

20. Sandars J, Lafferty N: **Twelve Tips on usability testing to develop effective e-learning in medical education.** *Medical teacher* 2010, **32**(12):956–60, [http://www.ncbi.nlm.nih.gov/pubmed/21090948].

21. Patel VL, Yoskowitz NA, Arocha JF, Shortliffe EH: **Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education.** *Journal of biomedical informatics* 2009, **42**:176–97, [http://dx.doi.org/10.1016/j.jbi.2008.12.002http://www.ncbi.nlm.nih.gov/pubmed/19135173].

22. Sweller J, van Merrienboer JJ, Paas FG: **Cognitive Architecture and Instructional Design**. *Educational Psychology Review* 1998, **10**(3):251–296, [http://doc.utwente.nl/58655/].

23. Sweller J: **Cognitive Load During Problem Solving: Effects on Learning**. *Cognitive Science* 1988, **12**(2):257–285, [http://doi.wiley.com/10.1207/s15516709cog1202_4].

24. Dror I, Schmidt P, O'connor L: **A cognitive perspective on technology enhanced learning in medical training: great opportunities, pitfalls and challenges.** *Medical teacher* 2011, **33**(4):291–6, [http://www.ncbi.nlm.nih.gov/pubmed/21456986].

25. van Merri_nboer JJG, Sweller J, Merrie JJGV: **Cognitive load theory in health professional education: design principles and strategies.** *Medical educationVan Merri_nboer, J. J. G., Sweller, J., & Merrie, J. J. G. Van. (2010). Cognitive load theory in health professional education: design principles and strategies. Medical education, 44(1), 85–93. doi:10.1111/j.1365-2923.2009.03498.x* 2010, **44**:85–93, [http://www.ncbi.nlm.nih.gov/pubmed/20078759].

26. Khalil MK, Paas F, Johnson TE, Payer AF: **Interactive and dynamic visualizations in teaching and learning of anatomy: a cognitive load perspective.** *Anatomical record. Part B, New anatomist* 2005, **286**:8–14, [http://www.ncbi.nlm.nih.gov/pubmed/16177993].

27. Sweller J: **Cognitive load theory, learning difficulty, and instructional design**. *Learning and instruction* 1994, **4**(4):295–312.

28. Kerfoot BP, Baker H: **An online spaced-education game to teach and assess residents: a multi-institutional prospective trial.** *Journal of the American College of Surgeons* 2012, **214**(3):367–73, [http://www.ncbi.nlm.nih.gov/pubmed/22225647].

29. Shaw T, Long A, Chopra S, Kerfoot BP: **Impact on clinical behavior of face-to-face continuing medical education blended with online spaced education: a randomized controlled trial.** *The Journal of continuing education in the health professions* 2011, **31**(2):103–8, [http://www.ncbi.nlm.nih.gov/pubmed/21671276].

30. Kerfoot BP, Fu Y, Baker H, Connelly D, Ritchey ML, Genega EM: **Online spaced education generates transfer and improves long-term retention of diagnostic skills: a randomized controlled trial.** *Journal of the American College of Surgeons* 2010, **211**(3):331–337.e1, [http://www.ncbi.nlm.nih.gov/pubmed/20800189].

31. Kerfoot BP, DeWolf WC, Masser BA, Church PA, Federman DD: **Spaced education improves the retention of clinical knowledge by medical students: a randomised controlled trial.** *Medical education* 2007, **41**:23–31, [http://www.ncbi.nlm.nih.gov/pubmed/17209889].

32. Kerdijk W, Tio RA, Mulder BF, Cohen-Schotanus J: **Cumulative assessment: strategic choices to influence students' study effort**. *BMC medical education* 2013, **13**:172.

33. Wood T: **Assessment not only drives learning, it may also help learning**. *Medical education* 2009, **43**:5–6.

34. Larsen DP, Butler AC, Roediger III HL: **Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial**. *Medical education* 2009, **43**(12):1174–1181.

35. Trial C, Cook DA, Thompson WG, Thomas KG, Thomas MR, Pankratz VS: **Impact of self-assessment questions and learning styles in Web-based learning: a randomized, controlled, crossover trial.** *Academic medicine : journal of the Association of American Medical Colleges* 2006, **81**(3):231–8, [http://www.ncbi.nlm.nih.gov/pubmed/16501263].

36. Taveira Gomes T, Severo M, Guimar_es J, Ferreira A: **A novel platoform for medical education - ALERT STUDENT**. *BMC Medical Education* 2014, **1**:1.

37. Ruiz JG, Mintzer MJ, Issenberg SB: **Learning objects in medical education.** *Medical teacher* 2006, **28**(7):599–605, [http://www.ncbi.nlm.nih.gov/pubmed/17594550].

38. Schwarz MR, Wojtczak A: **Global minimum essential requirements: a road towards competence-oriented medical education.** *Medical teacher* 2002, **24**(2):125–9, [http://www.ncbi.nlm.nih.gov/pubmed/12098430].

39. Landis JR, Koch GG: **The Measurement of Observer Agreement for Categorical Data**. *Biometrics* 1977, **33**:159–174.

40. Larsen DP, Butler AC, Roediger III HL: **Test-enhanced learning in medical education**. *Medical education* 2008, **42**(10):959–966.

41. Roediger HL, Karpicke JD: **The power of testing memory: Basic research and implications for educational practice**. *Perspectives on Psychological Science* 2006, **1**(3):181–210.

42. Schuwirth L, Vleuten Cvd, Donkers H: **A closer look at cueing effects in multiple-choice questions**. *Medical Education* 1996, **30**:44–49.

43. Schuwirth LW, Van Der Vleuten CP: **Different written assessment methods: what can be said about their strengths and weaknesses?** *Medical Education* 2004, **38**(9):974–979.

44. Cox M, Irby DM, Epstein RM: **Assessment in medical education**. *New England Journal of Medicine* 2007, **356**(4):387–396.

45. Sitzmann T, Ely K, Brown KG, Bauer KN: **Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure?** *The Academy of Management Learning and Education* 2010, **9**(2):169–191.

46. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L: **Accuracy of Physician Self-assessment Compared With Observed Measures of CompetenceA Systematic Review**. *Jama* 2006, **296**(9):1094–1102.

47. Tracey J, Arroll B, Barham P, Richmond D: **The validity of general practitioners' self assessment of knowledge: cross sectional study**. *BMJ* 1997, **315**(7120):1426–1428.

48. Harden RM, Gessner IH, Gunn M, Issenberg SB, Pringle SD, Stewart a: **Creating an e-learning module from learning objects using a commentary or 'personal learning**

assistant'. *Medical teacher* 2011, **33**(4):286–90, [http://www.ncbi.nlm.nih.gov/pubmed/21456985].

49. Clark R, Mayer R: **Applying the Segmenting and Pretraining Principles: Managing Complexity by Breaking a Lesson into Parts**. In *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia*, San Francisco: Jossey-Bass 2010:207–218.

50. Masters K, Ellaway R: **e-Learning in medical education Guide 32 Part 2: Technology, management and design.** *Medical teacher* 2008, **30**(5):474–89, [http://www.ncbi.nlm.nih.gov/pubmed/18576186].

51. Kim S, Song SM, Yoon YI: **Smart learning services based on smart cloud computing.** *Sensors (Basel, Switzerland)* 2011, **11**(8):7835–50, [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3231729\&tool=pmcentrez\&rendertype=abstract].

52. Martinez M: **Designing learning objects to personalize learning**. *PhD thesis*, Bloomington IL 2002.

53. Beux PL, Fieschi M: **Virtual biomedical universities and e-learning.** *International journal of medical informatics* 2007, **76**(5-6):331–5, [http://www.ncbi.nlm.nih.gov/pubmed/17407747].

54. Kristiansson MH, Troein M, Brorsson A: **We lived and breathed medicine-then life catches up: Medical students' reflections**. *BMC medical education* 2014, **14**:66.

Appendix A

Tables

Table A1: Study sample characterization

	Total	Control	Experiment	p
Gender	n (%)	n (%)	n (%)	
Female	59 (62.8)	28 (59.6)	31 (65.9)	0.670
Male	35 (37.2)	19 (40.4)	16 (34.1)	
Course year	n (%)	n (%)	n (%)	
4th year	44 (46.8)	23 (48.9)	21 (44.7)	0.836
5th year	50 (53.2)	24 (51.1)	26 (55.3)	
Preferred resource	n (%)	n (%)	n (%)	
Professor texts	36 (38.3)	17 (36.2)	19 (40.4)	0.898
Lecture notes	24 (25.5)	12 (25.5)	12 (25.5)	
Lecture slides	23 (24.5)	13 (27.7)	10 (21.3)	
Textbook	11 (11.7)	5 (11.6)	6 (12.8)	
Computer usage	n (%)	n (%)	n (%)	
Everyday	73 (77.7)	37 (78.2)	36 (76.6)	0.193
Not everyday	21 (22.3)	10 (21.2)	11 (23.4)	
Grades	Mean (SD)	Mean (SD)	Mean (SD)	
Cellular Biology	12.8 (1.2)	13.1 (1.6)	12.8 (1.6)	0.102
Course average	13.6 (1.1)	13.8 (1.1)	13.6 (1.1)	0.433
Daily study hours	Median (IR)	Median (IR)	Median (IR)	
During semester	3.0 (2.5)	3.0 (2.0)	3.0 (2.0)	0.628
During exam season	9.5 (2.0)	10.0 (2.0)	8.0 (2.0)	0.307

Table A2: Recall accuracy per session and group

	Total	Control	Experiment	
	Mean (SD)	Mean (SD)	Mean (SD)	p[1]
s0	0.76 (0.56)	0.72 (0.50)	0.81 (0.53)	0.924
s1	1.59 (0.67)	0.99 (0.54)	2.18 (0.55)	<0.001
s2	1.87 (0.65)	1.26 (0.62)	2.47 (0.45)	<0.001
p[2]	<0.001	<0.001	<0.001	<0.001[3]

SD - Standard Deviation; [1] Differences in recall accuracy between experiment and control group; [2] Differences in recall accuracy between pairwise sessions; [3] Interaction effect between session and group.

Table A3: Components of variance of recall accuracy for the control group

Component	n	Variance	SD	%[1]
Participant	47	0.165	0.406	15.1%
Flashcard	27	0.377	0.614	34.7%
Session	3	0.089	0.299	8.2%
Residual	3440	0.456	0.676	41.2%

SD - Standard Deviation; [1] Percentage of total variance.

Table A4: Components of variance of recall accuracy for the Experiment Group

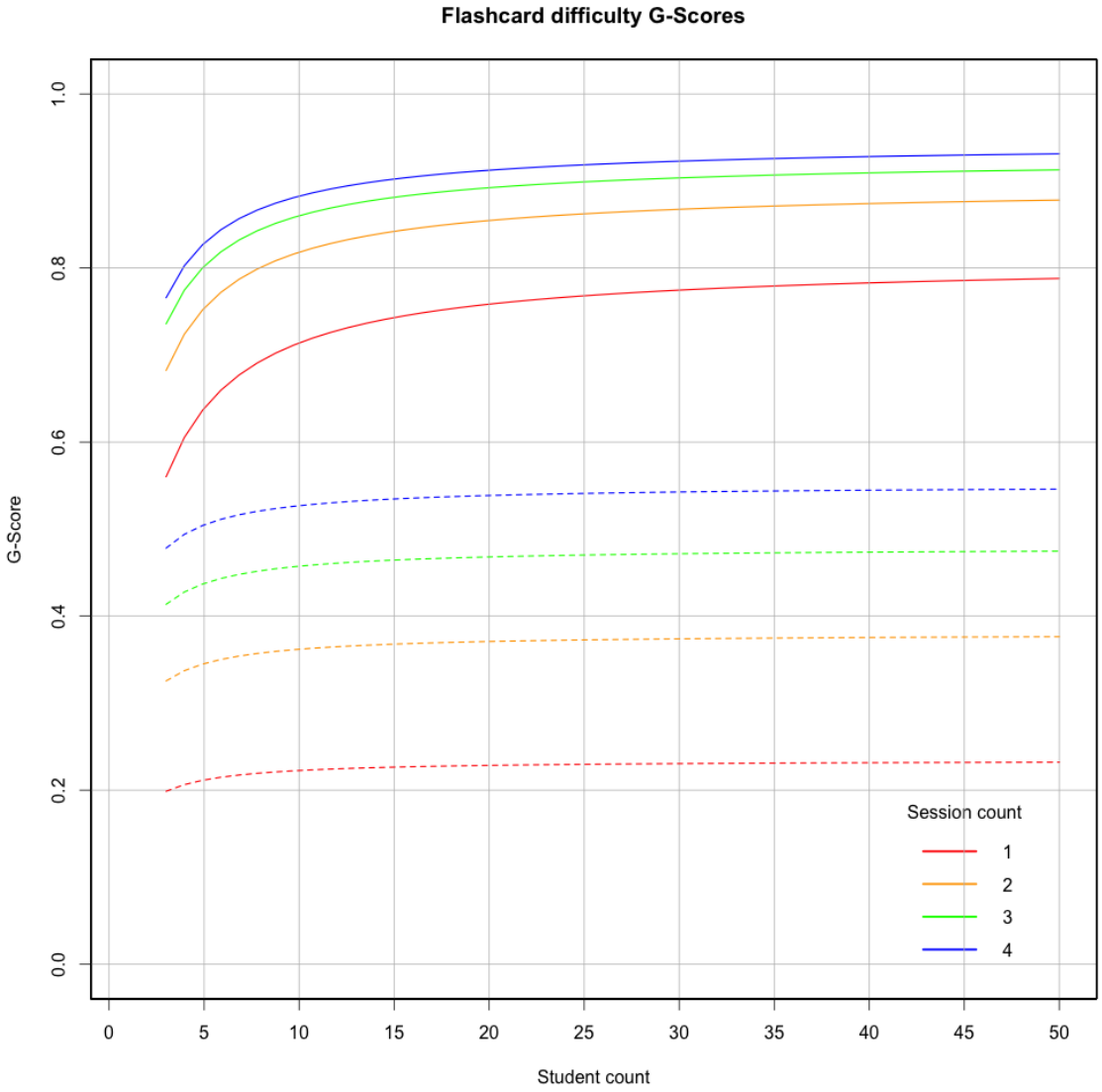
Component	n	Variance	SD	%[1]
Participant	47	0.083	0.288	5.1%
Flashcard	27	0.249	0.499	15.3%
Sessions	3	0.812	0.900	49.6%
Residual	3422	0.493	0.702	30.0%

SD - Standard Deviation; [1] Percentage of total variance.

Figures

Figure A1 - Flashcard recall-accuracy G-Score for each study group

G-Scores for the *flashcard* component of recall accuracy, using different combinations of number of students (x axis) and sessions (separate curves). The stroked line set represents control group scores, and the dashed line set represents experiment group scores. G-Scores between 0.81 - 1.00 indicate almost perfect agreement, 0.61 - 0.80 substantial agreement, 0.41 - 0.60 moderate agreement, 0.21 - 0.40 fair agreement, and values less than 0.21 depict poor or slight agreement.



Appendix B

Notebook original content (in Portuguese language)

Flashcard number	Flashcard content
1	O complexo de Golgi está envolvido no processamento e tráfego proteico (via de secreção) e também na síntese de glicolípidos e polissacarídeos. As proteínas são transportadas do retículo endoplasmático para a rede *cis* do Golgi e completam o processo de maturação na rede *trans* do Golgi onde as proteínas são empacotadas em vesículas para posterior transporte para os lisossomas (via endossomas), a membrana plasmática ou para o exterior da célula.
2	A designação de “aparelho de Golgi” é utilizada, por via de regra, para referir todos os complexos de Golgi de uma mesma célula.
3	A maturação das proteínas por N-glicosilação ocorre ao longo do transporte pelo complexo de Golgi.
4	No complexo de Golgi são sintetizados glicolípidos, esfíngomielina e outros polissacarídeos complexos que fazem parte da parede celular das plantas.
5	O complexo de Golgi é composto por um conjunto de cisternas (dictiossoma) e vesículas.
6	Distinguem-se 3 tipos de compartimentos funcionalmente distintos: • Rede *cis*• Cisternas (subdivididas em *cis*, medial e *trans*)• Rede *trans* do Golgi As vesículas provenientes do RE fusionam formando um compartimento intermédio entre o RE e Golgi, o ERGIC, de onde as proteínas são transportadas para a rede *cis* do Golgi.
7	As cisternas *cis*, medial e *trans* são os locais onde ocorre a maioria das reacções de processamento das proteínas. A rede *trans* do Golgi funciona como um centro de triagem e distribuição das proteínas para os endossomas, os lisossomas, a membrana plasmática ou o exterior da célula (secreção).
8	As proteínas do retículo endoplasmático (RE) entram via a face cis, também designada de formação, convexa e orientada para o núcleo na maioria das células.As proteínas são transportadas ao longo do complexo de Golgi e saem pela face trans, também designada de maturação, côncava, de onde as proteínas são direccionadas para os endossomas e lisossomas, a membrana plasmática e o exterior da célula (secreção), como ilustrado na figura.
9	As proteínas que devem permanecer no RE denominam-se proteínas residentes do RE, são transportadas de forma não específica do RE para o Golgi, e são recuperadas via transporte retrógrado para o RE. A recuperação de proteínas residentes do RE deve-se à presença de sinais de retenção na sua extremidade carboxilo.
10	Nas proteínas solúveis o sinal de retenção consiste em 4 aminoácidos, a sequência KDEL (Lys-Asp-Glu-Leu).Esta sequência liga-se especificamente a um receptor, o receptor KDEL, no ERGIC ou Golgi que permite o seu empacotamento em vesículas com revestimento COPI que efectuam transporte retrógrado para o RE.
11	Nas proteínas transmembranares o sinal de retenção consiste em 2 lisinas seguidas por 2 outros aminoácidos (KKXX) e liga-se directamente a vesículas revestidas com COPI que efectuam transporte retrógrado para o RE
12	As proteínas e lípidos provenientes do RE que têm por destino o complexo de Golgi são primeiro transportados para o ERGIC e seguidamente para a rede cis do Golgi através de vesículas de transporte com revestimento COPI.
13	Para além do processamento de glicoproteínas, o complexo de Golgi também está envolvido no metabolismo lipídico em particular na síntese de glicolípidos e esfingomielina.
14	A esfingomielina resulta da transferência de um grupo de fosforilcolina para a ceramida.
15	Os glicolípidos resultam da adição de hidratos de carbono à ceramida.

16	Nas plantas o complexo de Golgi está maioritariamente envolvido na síntese de polissacarídeos complexos que formam a parede celular.
17	Outro aspecto do processamento das glicoproteínas no complexo de Golgi consiste na adição de hidratos de carbono ao grupo OH de serina e treonina presentes em sequências peptídicas específicas (O-linked glicosilação).
18	A O-linked glicosilação é catalisada por uma série de glicosil transferases que adicionam primeiro N-acetilgalactosamina e de seguida um número variável de glícidos em geral até 10 glícidos.Em alguns casos estes glícidos serão ainda modificados pela adição de grupos sulfato.
19	Várias proteínas citosólicas e nucleares são processadas por O-glicosilação.
20	Nas proteínas destinadas à membrana plasmática, segundo passo ocorre nas cisternas mediais e consiste na remoção de mais 2 resíduos manose, e adição de 3 resíduos N-acetilglucosamina e fucose.
21	Um dos processos mais importantes na maturação das glicoproteínas nas cisternas do Golgi consiste na modificação dos oligossacarídeos N-linked, adicionados no RE, por uma sequência ordenada de reacções em cada cisterna.Nas proteínas destinadas à membrana plasmática ou secreção, a primeira modificação ocorre nas cisternas cis via a remoção de 3 resíduos manose.
22	Nas proteínas destinadas à membrana plasmática o último passo é completado nas cisternas trans através da adição de 3 resíduos galactose e por fim a adição de resíduos de ácido N-acetilneuramico a cada galactose.
23	O grau de processamento dos oligossacarídeos N-linked depende de:• Estrutura das proteínas nas cisternas do Golgi• Quantidade de enzimas nas cisternas do Golgi Em alguns casos as primeiras reacções de processamento (remoção de resíduos de manose) não ocorrem o que impede a adição posterior dos resíduos glícidicos, produzindo-se desta forma oligossacarídeos ricos em manose em vez de oligossacarideos complexos que seguem toda a via de processamento.
24	Nas proteínas destinadas aos lisossomas ocorre fosforilação de resíduos manose em duas reacções sequências.
25	Nas proteínas destinadas aos lisossomas a primeira reacção é catalisada na face cis pela enzima N-acetilglucosamina fosfotransferase.A N-acetilglucosamina fosfotransferase transfere um grupo N-acetilglucosamina fosfato para os resíduos de manose das hidroláses lisossomais.
26	A segunda reacção é catalisada por uma fosfodiesterase que remove o grupo N-acetilglucosamina deixando para trás o resíduo de manose fosforilado.
27	A especificidade do processamento das proteínas lisossomais reside na enzima N-acetilglucosamina fosfotransferase que catalisa a reacção de adição de N-acetilglucosamina fosfatos.Esta enzima reconhece um determinante estrutural presente unicamente nas proteínas lisossomais e denominado “signal patch” formado pela justaposição de sequências de aminoácidos provenientes de diferentes regiões da cadeia polipeptídica, como ilustrado na figura.

Notebook content – English translation

Flashcard number	Flashcard content
1	The Golgi complex is involved in protein processing, trafficking and the synthesis of glycolipids and polyssacharydes. The proteins are transported from the endoplasmic reticulum (ER) to the cis-Golgi network and complete the process of maturation in the trans-Golgi network, where the proteins are packed in vesicles for posterior transport to the lysossomes (via endossomes), the plasmatic membrane or to the cell exterior.
2	The designation "Golgi Apparatus" is used, as a general rule, to refer all the Golgi Complex in the same cell.
3	Protein maturation n-glycosilation occurs during the transport along the Golgi complex.
4	In the Golgi complex are synthesized glicolipids, sphyngomielin and other complex polyssacharides that make part of tha plant cell wall.
5	The Golgi complex is composed by a group of cisterns (dictiossomes) and vesicles.
6	There are 3 types of compartiments funcionally distinct: Cis network, Cisterns (subdivided in cis, medial and trans) Golgi network. The vesicles that come from the endoplasmatic retyculum fuse, forming an intermediate compartment between the RE and the Golgi, the ERGIC, where the proteins are transported to the Cis Golgi network.
7	The cis, medial and trans are the sites where the majority of the processing reactions occur. The trans Golgi network works as a center for triage and distribution of the proteins to the endossomes, the lisossomes the plasmatic membrane or the exterior of the cell.
8	The proteins of the endoplasmatic reticulum enter through the cis face, aslo designated as formation, convex and oriented towards the nucleus to the majority of the cells. The proteins are transporter along the Golgi complex, exit through the trans face, also designated maturation, concave, where the proteins are directed to the endossomes, lisossomes, the plasmatic membrane and the exterior of the cell, as illustrated in the picture.
9	The proteins that should remain in the endoplasmatic reticulum are named resident proteins on the endoplasmic reticulum, are transported in a non-specific manner from the ER to the Golgi, and are recuperated via a retrograde transport to the ER. The recuperation of proteins of the ER is owed to the presence of signals of retention in its carboxide extremity.
10	In the soluble proteins the retention sign consists of 4 aminoacids, the KDEL sequence (Lys-Asp-Glu-Leu). The sequence links specifically to a receptor, the KDEL receptor, on the ERGIC or Golgi, which allows its packaging in vesicles with COPI revestment that allow the retrogade transport to the ER.
11	In the transmembrane proteins the retention signal consists in 2 Lysine followed by other 2 other aminoacids (KXXX) and links directly to the vesicles revest with COPI that allow the retrograde transport to the RE.
12	The proteins and lipids that come from the ER that have the target the Golgi Complex are first transporter to the ERGIC and followed to the Cis network through the Cis Golgi network through the vesicles of transport with COPI revestment.
13	Besides the processing of glycoproteins, the Golgi Complex is also involved in the lipidic metabolism, in particular in the synthesis of glycolipids and sphingomielyn.
14	Sphyngomielin results in the transference a a phosphorincoline group to a ceramid molecule.
15	The glycoproteins result from the addition of carbohidrates to ceramide.
16	In the plants, the Golgi complex is majorly involved in the synthesis o polyssacharides that form the nuclear wall.
17	Other aspect of the processing of glycoproteins in the Golgi Complex consists in the addition of carbohydrates to the OH group on the resine and threonine residues present in specific peptidic sequences (O-linked glicosilation)

18	The O-linked glycosilation is catalized by a series of glycosil transferases that ass firstly a N-acetilgalactosamine and after a variable number of glicides, in general up to 10 residues. In some cases these residues are further modified by the addition of sulphate groups.
19	Some cytosolic and nuclear proteins are processes by O-glicosilation
20	In the proteins destined to the plasmaic membrane, the second step occurs in the medial cisterns and conssis in the remotion of more that 2 residues of manose, and the addition of 3 residues of N-acetilglucosamine and fucose.
21	One of the processes more important in the maturation of the glicoproteinsin the Golci cisterns consists in the modification of the N-linked oligossacharydes, added in the ER, by an ordered sequence of reactions in each cistern. In the proteins destined to the plasmatic membrane or secretion, the first modification occurs via remotion of 3 rediues of manose in the CIS cisterns.
22	In the proteins destined to the plasmatic membrane, the last step in completen in the trans cisterns through the addition of 3 rsidues of galactosis and by the addition of N-acetyl neuramic acid to each galactose molecule.
23	The degree of processin of the N-linked oligossacharides depends of: 1- The structure of the proteis ins the Golgi cisterns, 2. The quantuti of enzymes in the Golgi cisterns. Im some cases the first reactions of processins (remotion of mannose residues) do not occur, which preventes the following addition of glicidic residues, which leads to the formation of oligossacharydes rich in manose instead of complex oligossacharydes the follow the full processing pathway.
24	In the proteins destined to the lisossomes occurs the phosphorilation of manose residues in two sequenced reactions.
25	In the proteins destined to the lysossomes, the first reaction is catalised in the cis face by the enzyme n-acetilglucosamine phosphotransferase. The n-acetilglucosamine phosphotransferase transfers a group n-acetylglucosamine phosphate to the manose residues of the lisossomal hydrolases.
26	The second reaction is catalised by a phosphodiesterase that remoces the N-acetilglucosamine group, leaving behind a phosphorilated manose residue.
27	The specificity of the processing of the lysossomal proteins resides in the n-acetylglucosamine phosphotransferase enzyme, that catalyses the reaction of addition of N-acetylglucosamine phosphates. This enzyme recognizes a structural determinant present uniquely in the lisossomal proteins, named "signal patch", formed by the juxtaposition of aminoacid sequences from different regions of the polypeptidic chain, as illustrated in the picture.

Appendix C

Figure C1: Screenshot of *study mode* environment on ALERT STUDENT

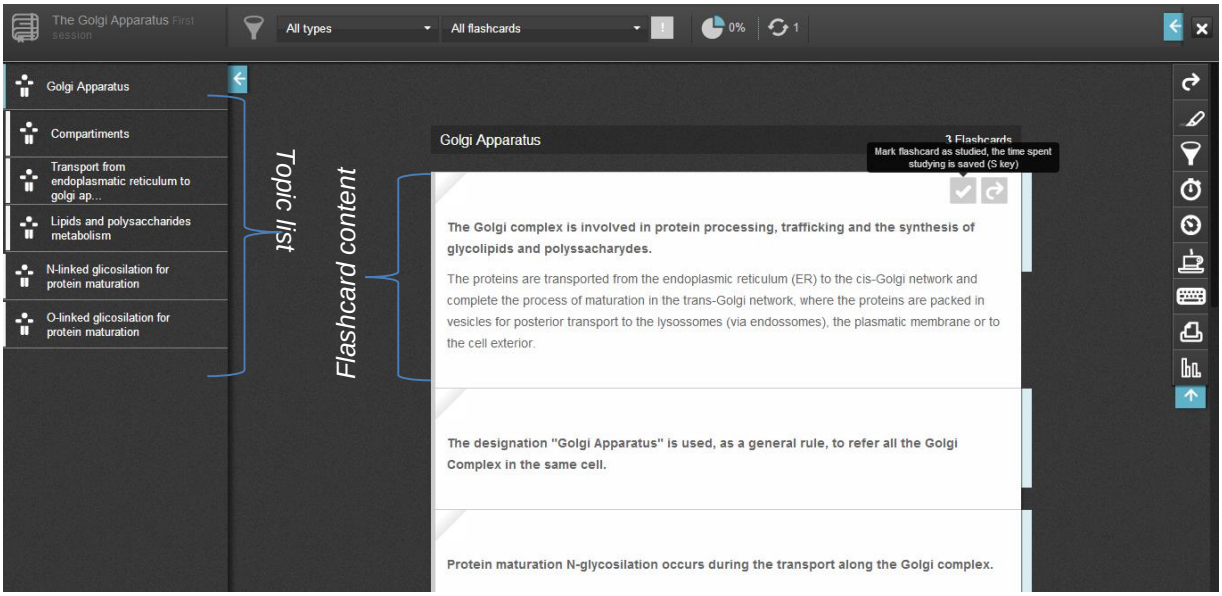


Figure C2: Screenshot of *quiz mode* environment on ALERT STUDENT